

# Assignment 2

Due: Oct 8th, 2018 at 1:59 pm

## Question 1: Geometric distribution (20 points)

Understand what is Geometric distribution and explain following functions:

1. Explain what is `dgeom()` and `pgeom()`. Create a simple question by yourself and answer it with these two functions
2. Explain what is `qgeom()`. Plug-in two numbers and explain the answer.
3. Explain what is `rgeom()`. Generate a sequence of result using this function and explain what is the meaning for them. (Requirement: show me at least 10 outputs)

## Question 2: Sample function (20 points)

Assuming there is a black box which contains 100 balls. 50 are red, 15 are blue, and remaining are yellow. Using `sample()` function to answer following questions:

1. Assume you get 1 ball out each time and put it back. What is the probability to see a yellow ball? Play this game for 10,000 times and calculate the probability based on the observation.
2. Assume you get 1 ball out until there is no balls anymore. What is the probability to see a red ball? Generate a plot to show me the dynamic of probability change based on the observation. Make sure your plot has the right x-label, y-label and legend.

## Question 3: Data cleaning and plot (30 points)

For this question, you need to download the “GOOGwNA.csv” file and answer following questions:

1. Removing all rows which has “NA” values. Then calculate the mean value for each columns using apply functions. (Except the date column)
2. Generate a plot using the high value column and low value column. In this plot, fill orange color between the low value and high value. Using gray color as background for data from July 1st, 2017 to Oct. 1st, 2017. Make sure your plot has the right x-label, y-label and legend.

## Question 4: Date and time (30 points)

For this question, you need to download the “BA.csv”. This is a sample data set from high frequency trading. It contains two type of data: “Trade” and “Quote”. Trade data shows the historical transaction price. Quote data shows the desired price from traders.

1. Use `subset()` function to get a sub-table which contains only “Trade” data.
2. Combine the “Date.L.” column and “Time.L.” column. Then, use `strptime()` to convert it into a time object.
3. Find out the last trade price in each minute. (e.x., from 9:30.000 to 9:30.999 you observe 200 records in total. The last record is the last trade price in this minute.)